DOCUMENT RESUME

ED 455 291 TM 033 145

AUTHOR Abedi, Jamal
TITLE Validity of Accommodation for English Language Learners.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 2001-04-00
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Dictionaries; *English (Second Language); *Junior High School Students; Junior High Schools; *Limited English Speaking; Science Instruction; Testing Problems; *Validity
IDENTIFIERS *Testing Accommodations (Limited Engl Proficiency)

ABSTRACT

The focus of this study was on the validity and feasibility of test accommodation strategies on a small-scale level. Both limited English proficiency (LEP) students and non-LEP students were tested under accommodated and nonaccommodated conditions and their performance was compared. The study was conducted in two public school districts and at one private school. A total of 422 students and 8 teachers from 6 school sites (14 eighth-grade science classes) participated. One form of accommodation consisted of English glosses and Spanish translations in the margins of the test booklet. The other form of accommodation consisted of a customized English language dictionary at the end of the test booklet. The dictionary contained only words used in the test items. The LEP students performed less well than the non-LEP students, and the difference was relatively large and statistically significant. The LEP students performed better under the accommodated conditions than under the standard condition. Accommodations had no significant effect on the scores of the non-LEP students. Results suggest that the customized dictionary enabled the LEP students to perform at a significantly higher level, with better results than for the glosses and translations. Results also show that the accommodation strategies did not impact the construct, and the validity of the assessment was not compromised. These results are encouraging given the ease of administration of these accommodations. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 455 291

VALIDITY OF ACCOMMODATION FOR ENGLISH LANGUAGE LEARNERS

Jamal Abedi

UCLA Graduate School of Education & Information Studies
Center for the Study of Evaluation
National Center for Research on Evaluation, Standards, and Student Testing

Paper presented at the 2001 Annual Meeting of the American Educational Research Association
Seattle, Washington

TM033145

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. Abedi
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The work reported herein was supported under the National Center of Education Statistics and Educational Research and Development Centers Program, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

Introduction

The literature on the assessment of students with limited English proficiency has found a significant link between students' language background and their performance in content-based areas. For example, studies by CRESST researchers have clearly demonstrated that language factors have significant impact of students' performance in math and science (Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Lord & Hofstetter, 1998, Abedi, Lord & Plummer, 1997). Following is a summary of some of major findings of CRESST studies:

1. When NAEP test items were grouped into long and short items, Abedi, Lord & Plummer (1997) found that LEP students performed significantly lower on the longer test items regardless of the level of content difficulty of the items. They also found that LEP students had higher proportions of omitted/not-reached items and had more difficulty with the items that were judged to be linguistically complex.
2. When math test items were modified to reduce the level of linguistic complexity, over 80% of middle-school students who were interviewed preferred the linguistically modified over the original English version of the test items (see Abedi et. al, 1997).
3. LEP students who received the modified English version of the math test items (approximately 700 students), performed significantly better than those receiving the original items (see Abedi et. al, 1997).
4. Spanish speaking students who received the Spanish translation of the NAEP math test (main assessment, 1996) performed significantly lower than the Spanish speaking students who received the English version of the test. We speculate that this is due to the impact of language of instruction on assessment (Abedi, Lord, and Hofstetter, 1998).
5. Consistent with the findings from the previous CRESST studies, among the three groups, LEP students who received the linguistically modified version of the tests (NAEP math items) performed the best, next were students receiving the original English version. As indicated above, students receiving

the Spanish version of the test, performed the lowest (Abedi, Lord, and Hofstetter, 1998).

6. Among the four accommodation strategies that were used (extra time, glossary, linguistically modified items and glossary plus extra time), the linguistically modified items was the only accommodation that reduced the performance gap between LEP and non-LEP students (Abedi, Hofstetter, Lord and Baker, 1998, 2000).

Studies that were summarized above clearly indicate that there is a substantial gap between performance of LEP and non-LEP and that this gap is mainly due to language factors. Previous studies have shown that utilizing some forms of language accommodations can increase test scores for LEP students and as a result can reduce the gap between performance of LEP and non-LEP students. For example, in an experimentally controlled study, Abedi, Hofstetter, Lord, and Baker (1998) found that a combination of glossary use and extra time increased LEP students' performance by over half a standard deviation. Other forms of accommodation, such as linguistic modification, may narrow the performance gap between LEP and non-LEP students (Abedi et al., 1997; Abedi, Hofstetter, Lord, and Baker, 1998).

Provision of accommodations has helped to increase the rate of inclusion for LEP students (Mazzeo, 1997). Based on the promising results, from using accommodations in the 1996 National Assessment for Educational Progress (NAEP) main assessment, accommodations were provided in the 1997 assessment in art and in the 1998 assessment in reading, writing, and civics.

There are, however, some major concerns regarding the use of accommodations for LEP students. Among the most important issues is the concern on the validity of accommodation strategies. As indicated earlier, providing accommodations has increased LEP students' performance, but at the same time non-LEP students have also benefited. This may be problematic, since the purpose of using accommodations is to reduce the gap between LEP and non-LEP students, not to alter the construct under measurement. The use of accommodation strategies, that affect the construct, is questionable.

The results of some of the CRESST studies have demonstrated that some forms of accommodations may impact the validity of assessment. Below are summaries of some of these studies:

1. Some forms of accommodation strategies such as glossary plus extra time raised performance of both LEP and non-LEP. The level of increase due to such accommodation strategies was higher for non-LEP students. This raised concern regarding the validity of accommodations (Abedi, Hofstetter, Lord and Baker, 1998, 2000).
2. English and bilingual dictionaries were used as different forms of accommodation strategies. The results of our studies suggested that by gaining access to definition of content-related terms, recipients of dictionary may be advantaged over those who did not have access to the dictionaries. This may jeopardize the validity of assessment (Abedi, Courtney, Mirocha, Leon and Goldberg, 2001).
3. The dictionary as a form of accommodation suffers from another major limitation, the feasibility issue. It was logistically very difficult to provide this form of accommodation to students (Abedi, Courtney, Mirocha, Leon and Goldberg, 2001).

The results of these studies clearly point to: (1) the impact of language factors in assessment, particularly for LEP students; (2) some forms of accommodation strategies help LEP students improving their performance and (3) some of the commonly used accommodation strategies may alter the construct under measurement.

A summary of one of our most recent studies in which new accommodation strategies were used and validity of accommodation is examined is given below as a sample of our CRESST studies.

Perspective

Recent federal and state legislation, including Goals 2000 and the Improving America's Schools Act (IASA), call for inclusion of all students in large-scale assessments such as the National Assessment for Educational Progress (NAEP). This includes students with limited English proficiency (LEP). However, we have clear evidence from recent research that students' language background factors impact their performance on content area assessments. For students with limited English proficiency, the language of the test item can be a barrier, preventing them from demonstrating their knowledge of the content area.

Various forms of testing accommodations have been proposed for LEP students. Empirical studies demonstrate that accommodation can increase test scores for both LEP and non-LEP students; furthermore, the provision of accommodations has helped to increase the rate of inclusion for LEP students in the NAEP and other large-scale assessments. There are, however, some major concerns regarding the use of accommodations for LEP students. Among the most important issues are those concerning the validity and feasibility of accommodation strategies.

7. Validity: The goal of accommodations is to level the playing field for LEP students, not to alter the construct under measurement. Consequently, if an accommodation affects the performance of non-LEP students, the validity of the accommodation could be questioned.

8. Feasibility: For an accommodation strategy to be useful, it must be implementable in large-scale assessments. Strategies that are expensive, impractical, or logistically complicated are unlikely to be widely accepted.

The focus of this study was on the validity and feasibility of accommodation strategies on small-scale level. In order to test for validity, both LEP and non-LEP students were tested under accommodated and non-accommodated conditions, and their performance was compared. Feasibility was a key consideration; we selected accommodation strategies for which

implementation would be practical in large-scale assessments. Since previous studies have identified the non-technical vocabulary of test items as a source of difficulty for LEP students (Abedi, Lord, and Plummer, 1995; Abedi, Hofstetter, and Lord, 1998); we chose two forms of accommodation targeting this issue.

Methodology

This study was conducted between November 1999 and February 2000, in two southern California school districts and at one private school site. The purpose of this study was to test the instruments, shed light on the issues concerning the administration of accommodations, explore the feasibility problems that we may encounter in other studies and, ultimately, provide data to help us modify the future study design. A total of 422 students and eight teachers, from six school sites (14 eighth-grade science classes), participated in this study.

A science test with 20 NAEP items was administered in three forms: one with the original items (no accommodation) and two with accommodations focusing on potentially difficult English vocabulary. One form of accommodation consisted of English glosses and Spanish translations in the margins of the test booklet. The other form of accommodation consisted of a customized English language dictionary at the end of the test booklet.

The customized dictionary - used in this study for the first time as an accommodation for LEP students - contained only words that are included in the test items. The customized English dictionary is grade appropriate and compiled by CRESST researchers. Providing full-length English dictionaries to test subjects has two major drawbacks: they are difficult to transport and they provide too much information on the content material being tested. For these reasons, the entries for non-technical words contained in the test have been excerpted (with permission from publisher) to create customized dictionaries that do not burden administrators and students with the bulk of a published dictionaries. Unlike the original dictionaries, these customized dictionaries do not contain words that assist the student with test content, thereby ensuring the validity of accommodations using dictionary. The pronunciation guide, font and type size are identical to that used in the original reference.

For each test booklet form, a follow-up questionnaire was developed to elicit student feedback. The Follow-up questionnaire was placed in the test booklet immediately after the science test. The questions were tailored to the type of science test the student completed. Students who received an accommodation were also asked if that accommodation helped them answer the science test items. Students' responses to these questions will be particularly helpful in designing the main study.

Included in the test booklet was the Science Background Questionnaire which included items selected from both the 1996 NAEP Grade 8 Bilingual Mathematics booklet and an earlier CRESST language background study. The questionnaire included queries regarding the student's country of origin, ethnicity, language background, language of instruction in science classes, and native language and English proficiency.

In their responses to the Science Background Questionnaire, most of the LEP students self-reported their ethnicity as Hispanic, followed by White, Asian, American Indian, and other. Most of the non-LEP students self-reported their ethnicity as White, followed by Hispanic, Asian, Black, American Indian, and other.

A science teacher questionnaire was also introduced midway through the study. This form was used at sites 4 through 6 to obtain information from each science teacher about each class, including type of science class, language of instruction; science topics covered so far this year, and students' English proficiency.

Test administrators received a science test administration script and were asked to complete a feedback questionnaire after each test administration. Test administrators distributed the six test booklets (three accommodation conditions by two forms) randomly within each classroom. The test directions were read aloud to the students. To address the different treatments, general directions were read aloud to the whole class, but specific directions were targeted to each treatment group. Students were given 25 minutes to complete the 20-item

science test, three minutes to complete the Follow-Up Questionnaire, and eight minutes to complete the Science Background Questionnaire.

Approval to conduct the study was received from The Office for Protection of Research Subjects (OPRS) at the University of California, Los Angeles (UCLA). Test administrators included CRESST research staff, retired teachers, and school administrators, who had prior experience with test administration. A letter to the principal described the study.

Results

This study examined the effectiveness of accommodations in addressing the difficulty of English vocabulary within test items in a NAEP science assessment. We compared LEP and non-LEP students' scores on 20 science items under three different conditions: customized dictionary, glossary, and standard NAEP condition (no accommodation). The analyses provided clear results with respect to the performance levels of LEP/non-LEP students, the effectiveness of the accommodations for LEP students, and the validity of the accommodated assessment.

4. Performance gap: LEP students performed lower than non-LEP students. For LEP students, the mean score was 8.97 (SD = 4.40, n=183) and for non-LEP students the mean was 11.66 (SD = 3.68, n=236). The difference between performance of LEP and non-LEP students is relatively large and is statistically significant ($t = 6.83$, $df = 417$, $p = .000$).

5. Effectiveness of accommodations: LEP students performed substantially higher under the accommodated conditions than under the standard condition. The mean for the LEP students under the customized dictionary was 10.18 (SD=5.26, n=55); under the glossary condition, the mean was 8.51 (SD=4.72, n=70); and under the standard condition the mean was 8.36 (SD=4.40, n=58). As the data suggest, LEP students did particularly well under the customized dictionary condition. The results of an analysis of variance (ANOVA) indicated that the difference between means for LEP

students under the three accommodation conditions was significant (F=3.08, df=2,180, p=.048).

6. Validity: The accommodations had no significant effect on the scores of the non-LEP students. For non-LEP students, the mean science score for the dictionary accommodation was 11.37 (SD=3.79, n=82); for the glossary the mean was 11.96 (SD=3.86, n=75); and for the standard condition the mean was 11.71 (SD=3.40, n=79). The results of analysis of variance showed no significant difference between the performance of non-LEP students under the three conditions (F=.495, df=2, 233, p=.610).

These results suggest that, first, the customized dictionary enabled LEP students to perform at a significantly higher level. Second, the accommodation strategies used in this study did not impact the construct, and the validity of the assessment was not compromised. These results are particularly encouraging, given the ease of administration of the accommodations that were used.

In student responses to the Follow-Up Questionnaires, LEP students reported greater difficulty with the language of the test items. (Follow-up questionnaires were similar but not identical for the three forms of the test.)

- More LEP than non-LEP students indicated there were words that they did not understand in the science test.
- LEP students, more than non-LEP students, wanted explanation of some of the difficult words.
- More LEP than non-LEP students expressed interest in using a dictionary during the test.
- LEP students, more than non-LEP students, indicated that it would have helped them if the test had explained words in another language.
- More LEP than non-LEP students expressed a preference for a dictionary during the test.

Analyses based on the background variables showed no significant gender differences. However, a significant difference was found between the performance of students who speak only English in the home and those who speak a language other than English in the home. Students who speak a language other than English performed significantly lower than the other group. This finding is consistent with the literature and with the main findings of this study.

Analyses of self-reported data showed that students who speak a language other than English in the home indicated that they speak that language more with their parents and less with their brothers, sisters, and friends. These findings, reflecting a generation gap, are consistent with the existing literature.

The results of analyses of self-reported data on English proficiency were also consistent with the literature and with the earlier findings of this study. As expected, LEP students reported significantly lower proficiency in English than their non-LEP counterparts.

Limitations

Since this was a pilot study and was planned to test the instruments and logistics for the main study, the generalizability of findings of this study is extremely limited. The generalizability of this study is further limited to grade level (Grade 8), content area (science), LEP language background (primarily Spanish), and accommodation type (dictionary and glossary).

It should also be noted that an accommodation for one grade level may not necessarily be appropriate, or even considered an accommodation, for another grade level. Students in lower elementary grades may not know how to use a dictionary or may be in the process of learning to use a dictionary, whereas students in higher elementary grade levels and above may be accustomed to regularly using a dictionary. For older students, dictionary use during a testing situation is considered an accommodation while for younger students dictionary may not be considered an effective form of accommodation since they may not know how to use it.

In an effort to find classrooms with an equal number of LEP and non-LEP students, site selection was based on state demographic information at the school site level. However, state demographic information does not necessarily reflect the LEP and non-LEP distribution for individual classes at a school site. Therefore, site selection in the main study should be based on demographic information collected at the classroom level.

A large proportion of the LEP population in southern California is native Spanish speaking. Accordingly, for the glossary accommodation we included English glosses and Spanish translations. In our sample, 88% of the LEP students were Hispanic and 26% of the non-LEP students were Hispanic. LEP students with first languages other than Spanish may have benefited from the English glosses, but the accommodation tells us little about the potential impact of translations in their first languages.

Implications and Recommendations

This study addresses several major issues concerning accommodations for LEP students in NAEP. Although these analyses report on the pilot phase of the study, there are nevertheless several implications for future NAEP assessments.

Since NAEP is a large-scale assessment, feasibility considerations are important. NAEP assessments involve a large number of LEP students, so ease of administration may be a determining factor. Any element that reduces the burden on states, schools, and students will potentially have a positive impact on future NAEP administrations. Educators are developing accommodation strategies that may reduce the gap between LEP and non-LEP scores in large-scale assessments. Not all of these strategies may turn out to be easily administered. One-on-one testing, for example, may be a highly effective form of accommodation, but it may not be feasible in large-scale assessments such as the NAEP.

Providing a customized dictionary is a viable alternative to providing traditional dictionaries. Dictionaries are, in fact, already widely used as instructional aids for LEP students, so the concept is not an unfamiliar one for students. Including a customized dictionary as part of the test booklet can

minimize the economic and administrative burden and may help to overcome shortcomings on the validity of accommodations using dictionaries. However, the economic and technical feasibility of providing a customized dictionary as a potential form of accommodation should be evaluated through cost-benefit analyses.

Gathering additional information about the academic performance and the language proficiency levels of students may help to clarify issues associated with inconsistency in the definition of LEP and the inclusion criteria for standardized assessments. The reading achievement data from Stanford 9, supplied by the schools, provided valuable information on the language proficiency levels of students, beyond the LEP designations. Given the inconsistency in the LEP designation criteria, collecting additional information about a student's academic and language performance would provide a more comprehensive picture of the student's academic knowledge. More accurate conclusions would be possible from analyses of contextual data, such as students' performance on other content areas and information on family and language background.

Abedi, J. Lord, C (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education*, 14,3, June 2001

Abedi, J. Lord, C. Hofstetter, C. & Baker, E. (2000) Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19, 3, pp. 16-26.

Abedi, J., Lord, C. & Hofstetter, C. (1998). Impact of Selected Background Variables on Students' NAEP Math Performance. Los Angeles: UCLA Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord C., & Plummer, J. R. (1997). *Final Report of Language Background as a Variable in NAEP Mathematics Performance.* Los Angeles: Center for the Study of Evaluation, CSE Technical Report # 429.

Anderson, N.E., Jenkins, F. F., & Miller, K.E. (1996). *NAEP Inclusion criteria and testing accommodations: Findings from the NAEP 1995 field test in mathematics.* Washington, D.C.: Educational Testing Service.

Anstrom, K. (1997). *Academic Achievement for Secondary Language Minority Students: Standards, Measures and Promising Practices.* Washington, DC.: National Clearinghouse for Bilingual Education.

Mazzeo, J. (1997, March). *Toward a more inclusive NAEP.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.